Remarks

Regarding amendments in the Specification

The terms "panel of markers" and "marker panel" have been added to the specification by amendment so that there is antecedent basis in the claims. The terms "panel of markers" and "marker panel" are well known in the art and have the same usage as the term "set of markers" in paragraph [018]. Applicants believe that this change is more reflective of the usage of these terms in the art and increases claim clarity.

Regarding amendments in the Claims:

The Examiner rejected claim 5 in the First Office Action dated August 23, 2004. Applicants have canceled claim 5.

Applicants are entitled to a total of 20 claims, including 3 independent claims subsequent to the RCE filed in the present application on March 4, 2004. Nineteen new claims, claims 6 through 24 inclusive (of which 2 claims are independent claims), have been added to the application and do not require additional claim fees.

Regarding new claim 6, this claim is essentially the same as pending, previously allowed claim 3 in related application No. 09/623, 068. The only difference being that the words "collection of points" in claim 3 has been changed to "collection of one or more points" in claim 6. This is supported by paragraph [0050]. If necessary to obviate a double patenting rejection, applicants will file a letter expressly abandoning application No. 09/623, 068. In addition, such action would allow the concentration of attention and economic resources in the present application. The subject matter of application No. 09/623, 068 is incorporated by reference into the present application in its entirety; and the present application claims priority from application No. 09/623, 068.

The Examiner rejected many of the claims in related application No. 09/623, 068 on the basis of failure to satisfy the written description requirement. In the event of traversal of the rejections, applicants were requested to indicate with particularity where in the specification support for limitations in the claims were to be found. (Applicants respectfully submit that at least some such support was cited in the Remarks section of the Amendment/Response of June 29, 2004 and in the two supplemental responses.) In order to expedite allowance of the presently pending claims in the present application, applicants will indicate with particularity support in the specification for limitations in the presently pending claims. This support will be cited in terms of paragraph numbers in square brackets. In addition, support is cited in the inventor's paper Annals of Human Genetics, 1998, vol 62, pp. 159-179, abbreviated AHG98 herein. This paper is incorporated by reference into the present patent application and a copy of AHG98 is also included herewith for the Examiner's convenience.

Applicants will cite support in the specification that would be apparent to a person of ordinary skill in the art. Regarding the written description requirement, applicants respectfully cite some quotes from the MPEP and case law. "[T]he 'essential goal' of the description of the invention requirement is to clearly convey the information that an applicant has invented the subject matter which is claimed." *In re Barker*, 559 F.2d 588, 592 n.4, 194 USPQ 470, 473 n.4 (CCPA 1977). To satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention. See, e.g., *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 USPQ2d at 1116.

What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail. See *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94. If a skilled-artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met. See, e.g., *Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972) (stating "the description need not be in *ipsis verbis* [i.e., "in the same words"] to be sufficient").

In addition, "Adequate description under the first paragraph of 35 USC 112 does not require literal support for the claimed invention...Rather, it is sufficient if the originally-filed disclosure would have reasonably conveyed to one having ordinary skill in the art that an appellant had possession of the concept of what is claimed." *Ex parte Parks*, 30 USPQ 2d 1234 (B.P.A.I. 1992).

Regarding new claim 7, the following support is cited. For the limitation *"wherein the CL-F region is for a species and a population and the population is a group of individuals as in the field of population genetics"*, e.g., see paragraphs [0175] bottom and [0135] (the term population is used in a statistical sense and in the sense the term population is used in the field of population genetics), e.g., see also [0306] the term population here is used as in the field of population genetics (e.g., Finnish populations). For the limitations *"wherein the CL-F region is N covered to within [x, y] by the two or more bi-allelic covering markers, wherein x is less than or equal to about D_{CL} or the equivalent thereof and y is less than or equal to about 0.2, D_{CL} is equal to the largest chromosomal length, computed by any method, for which linkage disequilibrium has been observed between any polymorphisms in any population of the species, N is an integer greater than or equal to 1"* e.g., see [0178], [0179], [0180], and [0080].

For the limitation in new claim 7: *"wherein the choice of covering markers is not based on the assumption that a covering marker is the trait-causing polymorphism"* e.g., see [0027] mid paragraph and [0029] mid to bottom paragraph (i.e. there is increased power even when the disease or trait causing allele is not the analyzed allele and m/p ratio departs from unity and, or linkage disequilibrium between the analyzed marker and trait-causing (disease) polymorphism is not maximal.) The same concepts are found throughout the application for example Table 2 p. 21 of the present application and Tables 1, 2 and 3 in AHG 98 pp. 165, and 167 all show increased power when the analyzed marker is not the trait-causing (or disease) polymorphism. (Disease is a genetic trait see [0059].) The Risch & Merikangas analysis is referred to in the Background of this patent application (e.g., see Risch & Merikangas analysis and the Muller-Myshok and Abel criticism/letter [0027]). In the Risch & Merikangas analysis, the TDT was assumed to test the disease locus itself or a perfectly associated bi-allelic marker, i.e. association with m = p and $\delta = \delta_{max}$ (e.g., see AHG 98 pp. 166 bottom, 169 top and Background [0027], [0029]); perfectly associated markers are also discussed on pp. 178, 179 of AHG 98 and [0316]). For each TDT association study or association test in the Risch & Merikangas analysis a marker allele with known association to each possible disease-causing polymorphism allele (in each such association study or test) is tested; and the known association values or data is m = p and $\delta = \delta_{max}$ for each such possible disease-causing polymorphism. The inventor's work has extended the analysis of Risch & Merikangas from the optimal situation of a perfectly associated marker for each possible disease-causing polymorphism (in each association study or test) to include the more common, less optimal situations in which $m \neq p$ and, or $\delta \neq \delta_{max}$ for tested marker-possible disease-causing polymorphism pairs (.e.g., see [0027, [0029] and final paragraph of AHG 98 pp. 170 bottom, 171 top.)

For the limitation in new claim 7: *"wherein the group of two or more covering markers is not an essentially one-dimensional panel of markers for a linkage study, wherein the essentially one-dimensional panel is a panel not based on using similarity of marker allele frequency and possible trait-causing polymorphism allele frequency to increase the power of an association-based linkage test to detect evidence for linkage"* see, e.g. [0019], [0020], top [0035] (i.e., conventional linkage study techniques are essentially one dimensional, focus on the dimension of chromosomal location but give little attention to the dimension of allele frequency) and see, e.g. [0308] "It is well known that increased disequilibrium between a marker and linked disease locus increases evidence for linkage provided by association-based linkage tests such as the TDT. However, what has not been recognized is that the specific allele frequencies of the marker locus can also have an enormous impact on the strength of evidence for linkage." And see a rendition of the principle that the inventor discovered: e.g. [0285] i.e., the power of association-based tests for linkage are increased as the allele frequencies of the disease-causing (or trait-causing) allele of a bi-allelic gene (or polymorphism) and a positively associated allele of a linked bi-allelic marker become similar in magnitude. That is, conventional (essentially one-dimensional) techniques are *not based on using similarity of marker allele frequency and possible trait-causing polymorphism allele frequency to increase the power of an association-based linkage test to detect evidence for linkage*. In addition, the application has been amended to include the well-known terms "panel of markers" and "marker panel" which have the same usage as the term "set of markers" in paragraph [0018].

For the record, the applicants note that the linkage disequilibrium in the well known principle quoted above from [0308] is essentially measured in a specific way: i.e. the increased disequilibrium is computed respectively as δ/δ_{max} for $\delta \geq 0$ or δ/δ_{min} for $\delta < 0$, wherein each of the δ values is a value of the coefficient of disequilibrium. This is the way (or essentially the way) that increased linkage disequilibrium is computed in the application in paragraphs [0291], [0292], in Table 2 on page 21, in AHG 98 in Tables1, 2, and 3 pp. 165, 167.

For support for new claim 8, see, e.g. [0075], [0050], [0090], that is CL-F regions may be large or small and a segment-subrange is a kind of CL-F region.

For support for new claim 9, support for the limitations beginning with *"wherein the CL-F region is for a species and a population"* and that ends with *"not based on using similarity of marker allele frequency and possible trait-causing polymorphism allele frequency to increase the power of an association-based linkage test to detect evidence for linkage"* has been cited above under new claim 7. For the remaining limitations in the claim, see, e.g. [0160], i.e. any method of systematically covering a CL-F region is acceptable. Such a method is taught in the Set/Subset Example paragraphs [0301] through [0321] inclusive of the Theory of Operation/Set/Subset Example [0281]. In this Set/Subset Example, a CL-F region is systematically covered using covering markers that are members of sets and subsets. Marker set and subset membership is based on the markers being located on particular chromosomal segments and having a particular least common allele frequencies. Each of the whereby clauses in this claim merely states the result of the invention recited in the claim and is not a limitation. As stated by the Federal Circuit, "[a] 'whereby' clause that merely states the result of limitations in the claim adds nothing to the patentability and substance of the claim." Texas Instruments, Inc. v. U.S. Int'l Trade Commission, 988 F 2d 1165, 1172, 26 USPQ2d 1018, 1023 (Fed. Cir. 1993).

For support for new claim 10, for the first limitation, see [0321], which states that Step 3 described in [0313] to [0317] is not essential (but it does increase efficiency). See, e.g. also [0271] which states that limiting the number of pairs of redundant covering markers is not crucial, but does increase efficiency. The application teaches the covering of a rectangular CL-F region using sets and subsets of covering markers; see, e.g. the Set/Subset Example paragraphs [0301] through [0321] cited supporting new claim 9 above, and see also, e.g., [0283] (marker least common allele frequencies vary systematically over a range or subrange and marker chromosomal locations vary systematically over one or more chromosomes or chromosomal regions), [0075] (A CL-F region may be large or small and can range over an entire chromosome or only a very small segment; and can range over the entire frequency range 0 to 0.5 or alternatively over only a very small subrange.); see also, e.g. [0321] which states that "versions of the invention are operable and have utility for <u>any subrange</u> of the least common allele frequency range 0 to 0.5" and [0185] (A segment-subrange is a rectangular CL-F region). The covered rectangular region is bounded by a chromosome or subregion of interest in the chromosomal location dimension and by a subrange in the allele frequency dimension. Each of the whereby clauses in this claim merely states the result of the invention recited in the claim and is not a limitation.

Regarding support for new claim 11, human being is a species described in the application.

Regarding support for new claim 12, see, e.g. [0321] which states that "versions of the invention are operable and have utility for any subrange of the least common allele frequency range 0 to 0.5" and see {0311] which describes covering the subrange "below 0.1/above 0.9" (i.e. the least common allele frequency subrange 0 to less than 0.1), see also, e.g. [0075] which says, "the least common allele frequency coordinates of CL-F points in a particular CL-F region can range over only a very small subrange" and gives an example of a subrange (the subrange 0.1 to 0.2) of small width (0.1). This width, 0.1, is the same as the width of the subrange 0 to less than 0.1.

Regarding support for new claim 13, "N ≥ 2" in claim 10 (from which claim 13 depends) literally means that N is greater than or equal to 2. As stated in [0182] "In general, the greater N is, the greater the power of a version of the invention for linkage studies". Thus the specification supports higher values of N, i.e. N>2.

Regarding support for new claim 14, see, e.g. [0324], which recites "thousands of bi-allelic markers". It was well-known in the art that large numbers of bi-allelic markers (e.g., thousands) would be available for use in linkage studies. For example, in the analysis of Risch and Merikangas [0027], 500,000 bi-allelic markers are studied. And the inventor's paper is a generalization of the Risch and Merikangas analysis [0029] (i.e. Risch and Merikangas is a special case of the inventor's general framework), see also e.g., [0249] which recites thousands of markers.

Regarding support for new claim 15, see, e.g. [0169], [0170], [0171]. See also, e.g. [0285] through [0289] inclusive and [0296] which describe increases in power along the allele frequency dimension, i.e. one or more gradients in power along the allele frequency dimension.

Regarding new claim 16, this claim is essentially the same as pending, previously allowed claim 78 in related application No. 09/623, 068. As stated above, if necessary to obviate a double patenting rejection, applicants will file a letter expressly abandoning application No. 09/623, 068. A difference being that the words "collection of points" in claim 78 has been changed to "collection of one or more points" in claim 16. This is supported by paragraph [0050].

The only other difference between claim 16 in the present application and claim 78 in the '068 application is that the words "substantially complementary" in claim 78 have been changed to simply "complementary" in claim 16. This brings claim 16 into verbatim conformity with the specification at, for example, [0265]. Claim 78 contains the words "substantially complementary". Given the definition of an oligonucleotide that is "complementary" [0141], the term "complementary" in this definition is essentially the same as "substantially complementary". And the applicants respectfully submit that the word "substantially" in "substantially complementary" of claim 78 is redundant. And the single word "complementary" means the same thing as "substantially complementary" in this context of claim 78. The applicants respectfully submit that a change from "substantially complementary" to "complementary" in claim 78 is a formality and does not change claim scope.

Regarding new claims 17 to 24, the limitations in these claims are very similar or essentially the same as those in new claims 7 to 15. Applicants have previously cited support in the specification for these limitations and the Examiner is respectfully referred to these comments above.

Conclusion

Single pending claim 5, rejected in the first Office Action post RCE, has been canceled. Nineteen new claims, claims 6 through 24 inclusive (of which 2 claims are independent claims), have been added to the application and do not require any additional claim fees. An appropriate payment of the fee for a one month extension is enclosed. The two new added independent claims, 6 and 16, are essentially the same as previously allowed claims in related application No. 09/623, 068. If necessary to obviate a double patenting rejection, applicants will file a letter expressly abandoning application No. 09/623, 068. Applicants have also cited parts of the present specification which support the new claims. (These citations of support are not necessarily exhaustive for the pending claims.)

For the reasons advanced above, applicants respectfully submit that the application is now in condition for allowance and that action is earnestly solicited.

Respectfully submitted,

Robert O. McGinnis

Robert O. McGinnis
Registration No. 44, 232

Dec. 23, 2004
1575 West Kagy Blvd.
Bozeman, MT. 59715
tel (406)-522-9355

Hidden linkage: a comparison of the affected sib pair (ASP) test and transmission/disequilibrium test (TDT)

R. E. McGINNIS
Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6145

(Received 3.11.97. Accepted 16.3.98)

SUMMARY

I compare the transmission/disequilibrium test (TDT) and affected sib pair (ASP) test under a general algebraic model describing a bi-allelic disease locus. Assuming linkage to a bi-allelic marker, I derive two binomial probabilities, one for parental allele 'transmission' (P_t) which determines the magnitude of the TDT χ^2 statistic (χ^2_{tdt}), and a second for identity-by-descent (ibd) marker allele 'sharing' (P_s) which determines the magnitude of the ASP test statistic (χ^2_{asp}). I also consider the ASP test applied to a completely polymorphic marker and demonstrate that the probability of ASP marker allele sharing (P_s) is identical to P_s observed for a bi-allelic marker in equilibrium with the disease locus. I present a general framework for determining the power of the TDT and ASP test based on expressions for P_t, P_s and the proportion (H/F) of ascertained parents who are informative at the marker. Two previous analytic investigations of TDT power based on the work of Ott (1989), and Risch & Merikangas (1996) are shown to be special cases of this general framework. In addition, I show the relationship between the framework I present and a third analytic investigation of TDT power for multi-allelic markers based on the work of Sham & Curtis (1995).

INTRODUCTION

Linkage has been demonstrated between insulin-dependent diabetes mellitus (IDDM) and the insulin gene region on chromosome 11p15.5 on the basis of linkage analysis by the transmission/disequilibrium test or TDT (McGinnis *et al.* 1991; Spielman *et al.* 1993). Linkage was demonstrated at the insulin 5′VNTR, a hypervariable marker that is extremely polymorphic, but whose VNTR alleles fall into two main size classes in Caucasians, thus forming a natural bi-allelic (+/−) marker. The + alleles were discovered to be positively associated with IDDM in case-control studies (Bell *et al.* 1984). Subsequent studies then demonstrated linkage in families collected for Genetic Analysis Workshop 5 (GAW5) by TDT analysis of GAW5 parents who were heterozygous (+/−) under the 5′VNTR bi-allelic categories (Spielman *et al.* 1993; see also Thomson *et al.* 1989, Julier *et al.* 1991).

The very strong evidence for linkage provided by the TDT ($\chi^2 = 8.26$, $p < 0.005$) was both surprising and puzzling because identity-by-descent (ibd) sharing of 5′VNTR alleles in affected sib pairs (ASPs) yielded no evidence for linkage in the same GAW5 families. Indeed, evidence for linkage was completely undetected or 'hidden' because the proportion of alleles shared by ASPs did not exceed the null hypothesis value of 0.5 in two different types of ASP analysis. On one hand, there was no increase in ASP allele sharing when the analysis included all GAW5 families in which both parents were informative for any two lengths of 5′VNTR allele (Spielman *et al.* 1989; Cox & Spielman, 1989). On the other hand, when the analysis included only those ASP parents who were evaluated by the TDT, namely those heterozygous (+/−) when the 5′VNTR is con-

Address for correspondence: Dr. Ralph McGinnis, Senior Investigator, SmithKline Beecham, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW.

sidered bi-allelic, there was again no evidence for linkage – in fact, fewer ASPs were concordant for a parental allele than were discordant (see table 7 in Spielman *et al.* 1993). Thus, whether the 5′VNTR was evaluated as highly polymorphic or as a bi-allelic marker associated with disease, ASP analysis failed to detect the strong evidence for linkage provided by the TDT.

This striking divergence in ASP and TDT linkage results illustrates that more information is needed about the relative power of the two tests under various conditions. Here I provide such information by analytically comparing the power of the TDT and ASP test as a function of variation in standard genetic parameters such as recombination fraction, disequilibrium, penetrance, and disease allele frequency. Based on a general algebraic model of linkage between a bi-allelic marker and bi-allelic disease locus, I derive two binomial probabilities. One probability for parental allele 'transmission' (P_t) determines the magnitude of the TDT χ^2 statistic (χ^2_{tdt}), and a second probability for ibd marker allele 'sharing' (P_s) determines the magnitude of the ASP test statistic (χ^2_{asp}). I also consider the ASP test applied to a completely polymorphic marker linked to a bi-allelic disease locus. In this situation, the probability of ASP marker allele sharing (P_s) is demonstrated to be identical to P_s observed for a bi-allelic marker in *equilibrium* with the disease locus.

The major findings of my investigation are as follows:

(1) The TDT, but not ASP test, can detect linkage with high power, when homozygotes for a susceptibility allele have only 2 to 4-fold greater disease risk than homozygotes for the normal or wild-type allele.

(2) TDT power is increased by disequilibrium between a bi-allelic marker and disease locus, and is also markedly increased when the disease allele and positively associated marker allele have similar population frequencies.

(3) The algebraic expressions for P_t and P_s each contain a product of three factors whose magnitude is added to the null hypothesis value of 0.5: $P_t = 0.5 + L_t M_t R_t$ and $P_s = 0.5 + L_s M_s R_s$ (see Results for the definition of each factor). The similarity of the corresponding factors in each expression facilitates ASP and TDT comparisons, and the different factors in each product enable some 'partitioning' of the contribution to evidence for linkage provided by different genetic parameters such as disequilibrium.

Based on the expressions for P_t, P_s and the proportion (H/F) of ascertained parents who are informative at the marker, I present a general framework for determining the power of the TDT and ASP test. Two previous analytic investigations of TDT power based on the work of Ott (1989), and Risch & Merikangas (1996) are shown to be special cases of this general framework. I also show the relationship between the framework I present and a third analytic investigation of TDT power for multi-allelic markers based on the work of Sham & Curtis (1995).

P_t and P_s determine the magnitudes of χ^2_{tdt} and χ^2_{asp}

Blackwelder & Elston (1985) investigated power to detect linkage for several varieties of ASP test. Among these tests, the t_2 'mean' test was found to be the most powerful ASP test for most genetic parameter values. Therefore, I chose the t_2 test (henceforth referred to simply as the 'ASP test') to compare with the power of the TDT. For nuclear families with at least two affected sibs, the ASP test considers each parent separately and determines whether the parent transmitted the same marker allele (ibd) to both sibs of an ASP. The well known χ^2 statistic for testing for linkage by the ASP test is:

$$\chi^2_{asp} = \frac{(n_s - n_u)^2}{(n_s + n_u)} = \frac{(n_s - n_u)^2}{n_{asp}},$$

where n_s and n_u are the number of instances in the data set in which a parental allele inherited by one affected sib is shared (n_s) or 'unshared' (n_u) by the second affected sib; thus $n_s+n_u = n_{asp}$ is the sample size for χ^2_{asp} and equals the number of trios in the data set consisting of an informative parent and an ASP.

Unlike the ASP test which usually considers ASP allele sharing from parents informative for any two marker alleles, the TDT only considers parents heterozygous for two particular marker alleles (e.g. A/B only). For a set of nuclear families, the TDT counts the number of times each A/B parent transmitted allele A or B to individual affected offspring. As shown by Spielman *et al.* (1993), the χ^2 statistic for detecting linkage by the TDT is:

$$\chi^2_{tdt} = \frac{(n_a-n_b)^2}{(n_a+n_b)} = \frac{(n_a-n_b)^2}{n_{tdt}}$$

where n_a and n_b are the number of instances in which an A/B parent transmitted allele A or B, respectively, to an individual affected offspring; and thus $n_a+n_b = n_{tdt}$ is the sample size for χ^2_{tdt}.

Note that the algebraic expressions for χ^2_{asp} and χ^2_{tdt} are identical in form. In each χ^2, the denominator is the sample size of the data set. Thus, when sample size (n_{asp} or n_{tdt}) is fixed, the denominator is constant and the magnitude of each χ^2 is determined only by the size of the squared difference in the numerator [$(n_s-n_u)^2$ or $(n_a-n_b)^2$].

A key idea in this paper is that the magnitude of the numerator in each χ^2 is determined by a specific binomial probability. In the case of χ^2_{asp}, this is the probability of ASP 'allele *sharing*' or P_s, i.e. the probability that a randomly ascertained parent of an ASP transmitted the same marker allele (ibd) to both affected sibs. In the absence of linkage, $P_s = 0.5$. But when linkage is present $P_s > 0.5$, and the larger the value of ($P_s-0.5$), the more ASPs that exhibit allele sharing (n_s) and the higher the magnitude of χ^2_{asp}. Similarly, a second binomial probability denoted P_t (for probability of 'allele *transmission*') determines the size of χ^2_{tdt}. P_t is the probability that marker allele A was transmitted to a specific affected child by a randomly ascertained A/B parent of an ASP. When linkage and disequilibrium are present, $P_t \neq 0.5$ and the larger the value of $|P_t-0.5|$, the greater the value of χ^2_{tdt}.

General algebraic model of linkage

At the beginning of Results, I give expressions for P_s and P_t based on the following general model: A bi-allelic marker with alleles A and B is linked to a bi-allelic disease locus with disease-predisposing allele D and non-predisposing allele d. The model allows any penetrance for the D/D, D/d and d/d genotypes (α, β, and γ, respectively) such that $1 \geqslant \alpha \geqslant 0$, $1 \geqslant \beta \geqslant 0$ and $1 \geqslant \gamma \geqslant 0$, and also assumes that no other locus underlies disease susceptibility. The recombination fraction (θ) between marker and disease locus is variable as are the population frequencies of the four marker-disease locus haplotypes [$f(AD) = c_1, f(Ad) = c_2, f(BD) = c_3, f(Bd) = c_4$, where $c_1+c_2+c_3+c_4 = 1$].

Note that once the haplotype frequencies are specified, the population frequency (p) of disease allele D is known ($p = c_1+c_3$), as are the frequencies ($m, 1-m$) of marker alleles A and B, respectively ($m = c_1+c_2; 1-m = c_3+c_4$). Furthermore, the coefficient of disequilibrium (δ) equals $c_1c_4-c_2c_3$ and thus, when convenient, the haplotype frequencies can be expressed as $c_1 = mp+\delta$, $c_2 = m(1-p)-\delta$, $c_3 = (1-m)p-\delta$, and $c_4 = (1-m)(1-p)+\delta$.

RESULTS

Based on derivations in Appendix I, equations (1) and (2) show expressions for P_s and P_t in terms of standard genetic variables for the general bi-allelic model described above. Both expressions assume that parents are ascertained through a randomly selected ASP, and each expression applies to an ascertained parent who is also heterozygous A/B at a bi-allelic marker. P_s is the probability that the parent transmitted the same marker allele to both affected sibs. P_t is the probability that the parent transmitted allele A to a particular affected child.

$$P_s = 0.5+(1-2\theta)^2 \; \frac{c_1 c_4+c_2 c_3}{H} \quad p^2\frac{(\alpha-\beta)^2}{4}+2p(1-p)\frac{(\alpha-\gamma)^2}{16}+(1-p)^2\frac{(\beta-\gamma)^2}{4}$$

Equation 1

$$P_t = 0.5+(1-2\theta) \; \frac{c_1 c_4-c_2 c_3}{H} \quad p^2 \; \frac{\alpha^2-\beta^2}{4} \; +2p(1-p) \; \frac{(\alpha+\beta)^2-(\beta+\gamma)^2}{16} \; +(1-p)^2 \; \frac{\beta^2-\gamma^2}{4}$$

Equation 2

Note that the expressions for P_s and P_t are similar in form. When there is no linkage, both expressions equal 0.5; but when linkage is present an amount is added to 0.5 which, in each expression, is a product of three factors. In both expressions, the leftmost factor depends on the recombination fraction (θ), the middle factor on haplotype frequencies (c_1, c_2, c_3, c_4) and the quantity H (see Appendix I), and the rightmost factor on penetrances (α, β, γ) and the frequency (p) of disease allele D. Because they play analogous roles in each expression, I denote the leftmost factor in P_s and P_t as L_s and L_t, respectively; and similarly denote the rightmost factor as R_s and R_t, and middle factor as M_s and M_t. Thus, $P_s = 0.5+L_s M_s R_s$ while $P_t = 0.5+L_t M_t R_t$.

Why $|P_t-0{\cdot}5| > (P_s-0{\cdot}5)$ *when disequilibrium is extreme*

As described in the Introduction, the ASP approach failed to detect linkage at the insulin 5′VNTR because the proportion of marker allele sharing in ASPs was close to 0.5, i.e. $P_s \approx 0.5$ whether the 5′VNTR was treated as bi-allelic or as highly polymorphic. By contrast, the TDT was able to detect linkage in the same families because $P_t \geqslant 0.60$ (see Spielman *et al.* 1993). Thus, regardless of differences in relative sample size ($n_{\text{asp}}, n_{\text{tdt}}$) for χ^2_{asp} and χ^2_{tdt}, the relative magnitudes of $(P_s-0.5)$ and $|P_t-0.5|$ are often the critical factor that causes a substantial difference in power for χ^2_{asp} and χ^2_{tdt}.

It is therefore interesting that analysis of equations (1) and (2) (see below) shows that when disequilibrium (δ) reaches its most positive value ($\delta_{\max}$) or its most negative value ($\delta_{\min}$), the magnitudes of P_s and P_t are such that: (a) $|P_t-0.5|$ and $(P_s-0.5)$ are both maximized and (b) $|P_t-0.5| > (P_s-0.5)$. This dependence on δ of $|P_t-0.5|$ and $(P_s-0.5)$ also has other important implications since the value of P_s for a *completely polymorphic* marker is identical to the P_s value of a bi-allelic marker in *equilibrium* ($\delta = 0$) with a bi-allelic disease locus (see Appendix IV). Therefore, if ASP allele sharing for a completely polymorphic marker is denoted by $(P_s-0.5)_{\delta=0}$, then P_t and P_s for any bi-allelic marker in extreme disequilibrium with the bi-allelic disease locus are such that: $|P_t-0.5| > (P_s-0.5) > (P_s-0.5)_{\delta=0}$.

To understand the pivotal role of δ in maximizing $|P_t-0.5|$ and $(P_s-0.5)$, and in determining their relative magnitudes, consider the three corresponding factors in P_t and P_s. By inspection, $L_t = (1-2\theta) \geqslant L_s = (1-2\theta)^2$, and when disequilibrium is present, θ should be near 0 and hence $L_t \approx L_s \approx 1$, the maximum value of each factor. Furthermore, as shown below, R_t is substantially greater than R_s, and the difference between the two factors is independent δ and θ since R_t and R_s depend only on the properties of the disease locus (p, α, β, γ). Therefore, $L_t \geqslant L_s$ and $R_t > R_s$, so it follows that $|P_t-0.5|$ would always exceed $(P_s-0.5)$ were it not for the influence of the remaining two factors in equations (1) and (2) (M_t and M_s).

Note, then, that M_t and M_s have the same denominator (H as defined in Appendix I); but the numerator of M_t is $\delta = c_1 c_4 - c_2 c_3$, while the numerator of M_s is the two components of δ *added* *together* $(c_1c_4+c_2c_3)$ implying that $M_s \geqslant |M_t|$. Since $|M_t|$ reaches its minimum value of 0 at equilibrium ($\delta = 0$), while M_s is always positive, it follows that $(P_s-0.5) > |P_t-0.5| \approx 0$ in an interval of δ values around $\delta = 0$. However, in Appendix III, I assume that marker allele frequency (m) and disease allele frequency (p) are fixed, and then show that $|M_t| = M_s$ at $\delta = \delta_{\max}$ and at $\delta = \delta_{\min}$. I also show that $|M_t|$ and M_s are both maximized at one of the two extreme δ values ($\delta_{\max}$ *or* $\delta_{\min}$). Therefore, for any bi-allelic marker (i.e. any m and p), when δ equals $\delta_{\max}$ or $\delta_{\min}$, $|P_t-0.5|$ and $(P_s-0.5)$ are maximized since $|M_t|$ and M_s are maximized; furthermore, because $|M_t| = M_s$ and $L_t \approx L_s$, the greater magnitude of R_t compared to R_s drives $|P_t-0.5|$ higher than $(P_s-0.5)$.

Why $R_t > R_s$

To understand why $R_t > R_s$, note first that both factors contain three components that are multiplied by the coefficients $p^2/4$, $[2p(1-p)]/16$ and $(1-p)^2/4$, respectively. In R_s, each component has the form $(U-V)^2$ while the corresponding component in R_t is (U^2-V^2) where U and V (for the components multiplied by $p^2/4$, $[2p(1-p)]/16$ and $(1-p)^2/4$) are, respectively, $U = \alpha, \alpha+\beta, \beta$ and $V = \beta, \beta+\gamma, \gamma$. Under the assumption that D/D penetrance exceeds d/d penetrance ($\alpha > \gamma$) and that D/d penetrance (β) lies somewhere between ($\alpha \geqslant \beta \geqslant \gamma$), each component in R_t $[U^2-V^2 = (U+V)(U-V)]$ must exceed its counterpart in R_s $[(U-V)^2]$ since $(U+V) > (U-V)$. The only exceptions occur when mode of inheritance is dominant ($\alpha = \beta$) or recessive ($\beta = \gamma$) in which case one pair of analogous components in R_t and R_s are equal; however, the other two components in R_t still exceed their counterparts in R_s, and thus $R_t > R_s$.

To assess how the elevation of R_t above R_s is influenced by the degree of risk conferred by the disease locus, the risk can be quantified by considering the penetrance of the D/D homozygote (α) to be r times greater than the penetrance of the d/d homozygote (γ). Thus, $\alpha = r\gamma$ and the penetrance of D/d (β) can be considered to fall between α and γ by letting $\beta = \gamma+x(\alpha-\gamma) = \gamma+x(r-1)\gamma$ where x is a number between 0 and 1. Based on this parameterization, the $\alpha{:}\gamma$ penetrance ratio (r) can be evaluated for its influence on R_t and R_s by dividing each component in R_t by its conterpart in R_s which yields the ratios:

$$\frac{1}{1-x}\;1+x+\frac{2}{r-1}\,,\quad 2\;\;x+\frac{1}{2}+\frac{2}{r-1}\,,\quad 1+\frac{1}{x}\;\frac{2}{r-1}\,.$$

Note that r appears in each ratio only in the term $2/(r-1)$ implying that R_t/R_s increases monotonically as r decreases. Thus, the elevation of R_t above R_s is most extreme for susceptibility loci causing a modest increase in disease risk as indicated by low values of r.

In Tables 1 and 2 below, I show values of P_s and P_t when $r = 2$ and $r = 4$, respectively. In these tables, $(P_s-0.5) \approx 0$ indicating that linkage would be difficult to detect by the ASP approach; but $|P_t-0.5|$ is much greater than $(P_s-0.5)$ when disequilibrium is extreme, thus illustrating that at low r values, R_t drives $|P_t-0.5|$ to levels that provide strong evidence for linkage.

Power of χ^2_{asp} *and* χ^2_{tdt}

I now show how to calculate and compare the power of χ^2_{asp} and χ^2_{tdt} when (a) both tests are applied to a bi-allelic marker or (b) the TDT is applied to a bi-allelic marker but the ASP test considers a completely polymorphic marker. I assume the two tests evaluate a series of S randomly ascertained parents of one or more ASPs, and that each test considers one ASP per parent. In the Discussion, I explain how to calculate the proportion (H/F) of the S parents who are informative at a bi-allelic marker. (The quantity F is proportional to the population frequency of parents who have two or more affected children, and H is proportional to the frequency of such parents who are also heterozygous at the marker.) Thus H/F determines the sample size for χ^2_{asp} and χ^2_{tdt} (see

Proportion of A/B parents in ascertained families). For instance, if both tests evaluate the same bi-allelic marker, then sample size for χ^2_{asp} is $n_{\text{asp}} = (H/F)S$, while sample size for χ^2_{tdt} is twice as large ($n_{\text{tdt}} = 2(H/F)S$) since χ^2_{asp} counts pairs of transmitted alleles while χ^2_{tdt} counts individual alleles. Based on these sample sizes ($n_{\text{asp}}, n_{\text{tdt}}$) and the values of P_s and P_t, the power of χ^2_{asp} and χ^2_{tdt} are determined from the binomial distributions

$$\frac{n_{\text{asp}}!}{n_s!n_u!}P_s^{n_s}(1-P_s)^{n_u} \quad \text{and} \quad \frac{n_{\text{tdt}}!}{n_a!n_b!}P_t^{n_a}(1-P_t)^{n_b},$$

respectively, as explained below.

Similarly, the power of χ^2_{asp} when applied to a completely polymorphic marker can also be determined from the appropriate binomial distribution, but $n_{\text{asp}} = S$ since all parents are informative, and $P_s = 0.5+(1-2\theta)^2[p(1-p)/F]R_s$ as shown in Appendix IV. Interestingly, this expression for P_s when the marker is completely polymorphic is identical to P_s for a bi-allelic marker in equilibrium with a bi-allelic disease locus. This can be verified by setting $\delta = 0$ in $c_1 = mp+\delta$, $c_2 = m(1-p)-\delta$, $c_3 = (1-m)p-\delta$, $c_4 = (1-m)(1-p)+\delta$ and substituting for the four haplotype frequencies in the expression for H (see Appendix I) and in equation (1).

Based on sample size ($n_{\text{asp}}, n_{\text{tdt}}$) and binomial probability (P_s, P_t), two binomial distributions are generated which can be used to calculate the power of χ^2_{asp} and χ^2_{tdt} as described in Appendix II. Specifically, the power of χ^2_{asp} or the probability that $\chi^2_{\text{asp}} > L$ (a significance cutpoint) is equal to the portion of the binomial distribution

$$\frac{n_{\text{asp}}!}{n_s!n_u!}P_s^{n_s}(1-P_s)^{n_u} \quad \text{for which} \quad n_s > \frac{n_{\text{asp}}}{2}+\frac{\surd(n_{\text{asp}}L)}{2}.$$

Similarly, if marker allele A is associated with disease, the power of χ^2_{tdt} is estimated by the portion of the binomial distribution

$$\frac{n_{\text{tdt}}!}{n_a!n_b!}P_t^{n_a}(1-P_t)^{n_b} \quad \text{for which} \quad n_a > \frac{n_{\text{tdt}}}{2}+\frac{\surd(n_{\text{tdt}}L)}{2}.$$

Thus, standard tables giving the normal approximation to the binomial distribution (Pearson & Hartley, 1954; Weir, 1996) provide precise power values for virtually any sample size ($n_{\text{asp}}, n_{\text{tdt}}$), binomial probability (P_s, P_t), and significance level.

Comparison of TDT and ASP power

Here I illustrate how the equations for P_t, P_s and H/F can be used to compare the power of χ^2_{tdt} and χ^2_{asp}. I assume the two tests consider markers that are tightly linked ($\theta = 0$) to bi-allelic disease loci with additive mode of inheritance ($\beta = (\alpha+\gamma)/2$) and for which the $\alpha{:}\gamma$ penetrance ratio is $r = 2$, $r = 4$ or $r = 10$. Penetrance ratios of $r = 2$, 4 and 10 were chosen as being somewhat representative of the entire genetic parameter space since I have found that P_t and P_s increase rapidly as r increases from 2 to 6 with smaller, asymptotic increases in P_t and P_s for $r > 10$. Furthermore, additive mode of inheritance may also be regarded as being somewhat representative since results from other modes of inheritance do not, in general, substantially differ from results presented here. In the tables below, I compare χ^2_{tdt} and χ^2_{asp} when both tests consider the same bi-allelic marker, or when χ^2_{asp} considers a fully informative marker and χ^2_{tdt} evaluates a nearby bi-allelic marker. Such single test comparisons would be occasioned by: (a) TDT and ASP analysis of a marker that gave 'suggestive' evidence of linkage and disease-association in other families or in comparisons of allele frequencies in cases and unrelated controls; or (b) TDT and ASP analysis of markers near a candidate gene suspected of increasing disease susceptibility.

Table 1. *ASP and TDT power for α:γ penetrance ratio of $r = 2$*[a]

					TDT and ASP test of the same bi-allelic marker									
	ASP test of fully informative marker[c]				$\delta = \delta_{\max}$[b]					$\delta = \frac{1}{2}\delta_{\max}$[b]				
			Power					Power					Power	
dis. allele freq(p)[b]	P_s	H/F	ASP	mkr. allele freq(m)[b]	P_t	P_s	H/F	TDT	ASP	P_t	P_s	H/F	TDT	ASP
$p = 0.60$	0.506	1.0	0.08	$m = 0.75$	0.568	0.511	0.33	0.60	0.08	0.532	0.507	0.35	0.19	0.07
				$m = 0.50$	0.560	0.510	0.50	0.67	0.09	0.530	0.507	0.50	0.22	0.07
				$m = 0.25$	0.537	0.506	0.40	0.26	0.07	0.519	0.506	0.39	0.10	0.07
$p = 0.40$	0.507	1.0	0.09	$m = 0.75$	0.553	0.509	0.34	0.42	0.07	0.525	0.508	0.36	0.13	0.07
				$m = 0.50$	0.573	0.513	0.50	0.83	0.10	0.536	0.509	0.50	0.30	0.08
				$m = 0.25$	0.563	0.511	0.43	0.65	0.09	0.534	0.508	0.40	0.23	0.07
$p = 0.15$	0.506	1.0	0.08	$m = 0.75$	0.526	0.505	0.36	0.14	0.06	0.513	0.505	0.37	0.07	0.07
				$m = 0.50$	0.537	0.507	0.50	0.32	0.07	0.518	0.506	0.50	0.11	0.07
				$m = 0.25$	0.565	0.512	0.42	0.67	0.09	0.534	0.509	0.40	0.23	0.08

[a] ASP power (1-tailed test) and TDT power (2-tailed test) for a significance level of 0.05 and sample size of 200 families; thus $n_{\text{asp}} = 400\ H/F$ and $n_{\text{tdt}} = 800\ H/F$.
[b] $\delta_{\max}$ ($\delta_{\min}$) is most positive (most negative) value of disequilibrium for bi-allelic marker and disease locus with allele frequencies m and p, respectively; power results shown for $\delta_{\max}$ ($1/2\delta_{\max}$) at $m = 0.75$, 0.5 and 0.25 equal power results for $\delta_{\min}$ ($1/2\delta_{\min}$) when $m = 0.25$, 0.5 and 0.75, respectively.
[c] P_s for a fully informative marker is identical to P_s for a bi-allelic marker at $\delta = 0$.

Table 2. *ASP and TDT power for α:γ penetrance ratio of $r = 4$*[a]

					TDT and ASP test of the same bi-allelic marker									
	ASP test of fully informative marker[c]				$\delta = \delta_{\max}$[b]					$\delta = \frac{1}{2}\delta_{\max}$[b]				
			Power					Power					Power	
dis. allele freq(p)[b]	P_s	H/F	ASP	mkr. allele freq(m)[b]	P_t	P_s	H/F	TDT	ASP	P_t	P_s	H/F	TDT	ASP
$p = 0.60$	0.516	1.0	0.16	$m = 0.75$	0.620	0.534	0.30	0.96	0.18	0.553	0.522	0.34	0.42	0.13
				$m = 0.50$	0.597	0.527	0.49	0.97	0.19	0.548	0.519	0.50	0.52	0.13
				$m = 0.25$	0.556	0.516	0.42	0.54	0.11	0.530	0.515	0.40	0.19	0.10
$p = 0.40$	0.525	1.0	0.26	$m = 0.75$	0.607	0.534	0.31	0.93	0.19	0.548	0.527	0.34	0.35	0.15
				$m = 0.50$	0.636	0.543	0.48	0.99	0.32	0.566	0.530	0.50	0.75	0.21
				$m = 0.25$	0.607	0.534	0.46	0.99	0.24	0.558	0.527	0.42	0.57	0.17
$p = 0.15$	0.530	1.0	0.33	$m = 0.75$	0.570	0.526	0.33	0.63	0.15	0.533	0.528	0.35	0.19	0.16
				$m = 0.50$	0.594	0.536	0.49	0.96	0.26	0.547	0.531	0.50	0.53	0.22
				$m = 0.25$	0.644	0.555	0.48	0.99	0.45	0.581	0.540	0.43	0.86	0.27

[a] ASP power (1-tailed test) and TDT power (2-tailed test) for a significance level of 0.05 and sample size of 200 families; thus $n_{\text{asp}} = 400\ H/F$ and $n_{\text{tdt}} = 800\ H/F$.
[b] $\delta_{\max}$ ($\delta_{\min}$) is most positive (most negative) value of disequilibrium for bi-allelic marker and disease locus with allele frequencies m and p, respectively; power results shown for $\delta_{\max}$ ($1/2\delta_{\max}$) at $m = 0.75$, 0.5 and 0.25 equal power results for $\delta_{\min}$ ($1/2\delta_{\min}$) when $m = 0.25$, 0.5 and 0.75, respectively.
[c] P_s for a fully informative marker is identical to P_s for a bi-allelic marker at $\delta = 0$.

In Table 1 ($r = 2$), Table 2 ($r = 4$) and Table 3 ($r = 10$), column 1 shows disease allele frequency (p) for a bi-allelic disease locus, and columns 2, 3 and 4 show results for χ^2_{asp} applied to a fully informative marker. The remaining columns in each table list χ^2_{tdt} and χ^2_{asp} results for a linked bi-allelic marker whose allele frequency (m) is listed in column 5. Results are given for each value of m (0.75, 0.50, 0.25) assuming positive disequilibrium between the bi-allelic marker allele and disease allele is maximal ($\delta = \delta_{\mathrm{max}}$) or half-maximal ($\delta = \frac{1}{2}\delta_{\mathrm{max}}$) where $\delta_{\mathrm{max}} = \min[(1-m)p, (1-p)m]$. TDT power (two-tailed test) and ASP power (one-tailed test) are for a significance level of 0.05 and are based on a sample size of 200 families (i.e. 400 parent-ASP trios) and thus $n_{\mathrm{asp}} = 400(H/F)$ while $n_{\mathrm{tdt}} = 800(H/F)$.

The ASP test can detect linkage over long distances ($\theta \gg 0$) and in the absence of disequilibrium ($\delta = 0$); but the TDT has no power when $\delta = 0$ and hence can detect linkage only over short distances (generally less than 1 cM). Yet when disequilibrium is half-maximal or greater ($\delta \geqslant \frac{1}{2}\delta_{\mathrm{max}}$), Tables 1, 2 and 3 each show that TDT power almost always exceeds ASP power whether χ^2_{asp} is applied to a fully informative or bi-allelic marker. When $r = 2$ (Table 1), linkage is virtually undetectable by χ^2_{asp} since $(P_s - 0.5) \leqslant 0.13$ and ASP power is 0.10 or lower; by contrast, the TDT is able to detect linkage but TDT power exceeds 0.50 only when δ is close to δ_{max} and allele frequencies (m, p) are similar in magnitude at the marker and disease locus. For $r = 4$ (Table 2), ASP power is increased but still relatively low ($\leqslant 0.33$) for fully informative markers and for most bi-allelic markers. TDT power is also substantially higher and, for most markers, exceeds 0.95 when $\delta = \delta_{\mathrm{max}}$ and exceeds 0.50 when $\delta = \frac{1}{2}\delta_{\mathrm{max}}$, thus indicating that when $r = 4$, the TDT could demonstrate linkage to many disease loci whose linkage might be difficult or impossible to establish by the ASP test.

For $r = 10$ (Table 3), TDT power is reasonably high ($\geqslant 0.66$) when $\delta \geqslant \frac{1}{2}\delta_{\mathrm{max}}$ and ASP power is also elevated (> 0.50) except at the highest disease allele frequency shown ($p = 0.6$) where ASP power is 0.28 for a fully informative marker. Thus as r increases from 2 to 10, the tables show that P_s and ASP power increase substantially and hence, when $r = 10$, the relative power advantage of the TDT is diminished. Nevertheless, as indicated by lower ASP power when $p = 0.6$ at $r = 10$ (Table 3), ASP power at elevated disease allele frequencies ($p > 0.6$) remains low (< 0.50) even when $r \rightarrow \infty$ (data not shown; table available from the author). For example, if the same power analysis shown in Tables 1–3 were conducted for $r = \infty$ (i.e. $\gamma = 0$) then for a fully informative marker and disease allele frequency of $p = 0.75$, P_s and ASP power would be 0.519 and 0.19, respectively. By contrast, TDT power would be much higher ($\geqslant 0.80$) but only when $\delta \geqslant \frac{3}{4}\delta_{\mathrm{max}}$ and m is close to $p = 0.75$ (i.e. $0.65 \leqslant m \leqslant 0.85$).

In concluding this section, I emphasize that Tables 1–3 show that when the disease locus and marker are bi-allelic, TDT power is substantially increased if the disease allele and positively associated marker allele have similar frequencies. Müller-Myshok & Abel (1997) independently made a similar observation, but they emphasized the weakness of TDT power when the m/p ratio departs from unity and δ is not close to δ_{max}. However, the tables illustrate that similar frequencies for the disease allele and associated marker allele can increase TDT power to reasonably high levels even when the m/p ratio substantially differs from 1 and δ is much lower than δ_{max}. For example, in Table 3 ($r = 4$), note that when $\delta = \frac{1}{2}\delta_{\mathrm{max}}$ and $p = 0.15$, a similar frequency ($m = 0.25$) for the disease-associated marker allele produces TDT power of 0.86 and P_t of 0.581; but when $p = 0.15$ and $m = 0.5$ at $\delta = \frac{1}{2}\delta_{\mathrm{max}}$, TDT power and P_t fall to 0.53 and 0.547, respectively. The difference in TDT power for these two situations can also be quantified by calculating the mean value of χ^2_{tdt} based on a sample of 200 ASP families and the values of P_t and H/F in Table 4 [i.e. $\chi^2_{\mathrm{tdt}} = 800(H/F)(2P_t - 1)^2$]. When $p = 0.15$ and $m = 0.5$, $\chi^2_{\mathrm{tdt}} = 3.53$ yielding a significance level of $p = 0.06$; but when $p = 0.15$ and $m = 0.25$, $\chi^2_{\mathrm{tdt}} = 9.02$ for a significance level of $p < 0.003$. The large

Table 3. *ASP and TDT power for $\alpha{:}\gamma$ penetrance ratio of $r = 10$*[a]

					TDT and ASP test of the same bi-allelic marker									
	ASP test of fully informative marker[c]				$\delta = \delta_{\max}$[b]					$\delta = \frac{1}{2}\delta_{\max}$[b]				
			Power					Power					Power	
dis. allele freq(p)[b]	P_s	H/F	ASP	mkr. allele freq(m)[b]	P_t	P_s	H/F	TDT	ASP	P_t	P_s	H/F	TDT	ASP
$p = 0.60$	0.527	1.0	0.28	$m = 0.75$	0.660	0.560	0.27	0.99	0.34	0.567	0.537	0.33	0.59	0.21
				$m = 0.50$	0.621	0.546	0.48	0.99	0.36	0.559	0.531	0.50	0.66	0.22
				$m = 0.25$	0.568	0.526	0.43	0.72	0.17	0.536	0.525	0.40	0.25	0.16
$p = 0.40$	0.547	1.0	0.59	$m = 0.75$	0.654	0.569	0.28	0.99	0.42	0.566	0.552	0.33	0.56	0.33
				$m = 0.50$	0.686	0.583	0.47	0.99	0.74	0.589	0.555	0.49	0.94	0.46
				$m = 0.25$	0.636	0.560	0.48	0.99	0.51	0.576	0.549	0.43	0.81	0.36
$p = 0.15$	0.580	1.0	0.94	$m = 0.75$	0.633	0.577	0.31	0.99	0.53	0.560	0.577	0.34	0.51	0.56
				$m = 0.50$	0.668	0.596	0.49	0.99	0.86	0.582	0.584	0.50	0.91	0.77
				$m = 0.25$	0.727	0.630	0.54	0.99	0.99	0.632	0.599	0.47	0.99	0.86

[a] ASP power (1-tailed test) and TDT power (2-tailed test) for a significance level of 0.05 and sample size of 200 families; thus $n_{\mathrm{asp}} = 400\ H/F$ and $n_{\mathrm{tdt}} = 800\ H/F$.

[b] $\delta_{\max}$ ($\delta_{\min}$) is most positive (most negative) value of disequilibrium for bi-allelic marker and disease locus with allele frequencies m and p, respectively; power results shown for $\delta_{\max}$ ($1/2\delta_{\max}$) at $m = 0.75$, 0.5 and 0.25 equal power results for $\delta_{\min}$ ($1/2\delta_{\min}$) when $m = 0.25$, 0.5 and 0.75, respectively.

[c] P_s for a fully informative marker is identical to P_s for a bi-allelic marker at $\delta = 0$.

difference in significance level (0.06 versus 0.003) and power (0.53 versus 0.86) illustrated by this example indicates that careful attention to allele frequencies at bi-allelic markers may play an important role in future efforts to map susceptibility loci.

DISCUSSION

The equations for P_s, P_t and H/F enable comparison of TDT and ASP power for the same family data, since the three expressions assume random ascertainment of parents of *two* or more affected children. However, the TDT can be applied to families with a single affected child, so in Appendix I, I derive an expression analogous to P_t (denoted P_t^*) which gives the probability that allele A was transmitted to an affected child by a randomly ascertained A/B parent of *one* or more affected offspring. The derivation of P_t^* is almost identical to that of P_t, and hence the algebraic form of P_t^* is similar to that of P_t and P_s:

$$P_t^* = 0.5+(1-2\theta)\ \frac{c_1 c_4 - c_2 c_3}{H^*}\quad p^2\frac{(\alpha-\beta)}{2}+p(1-p)\frac{(\alpha-\gamma)}{2}+(1-p)^2\frac{(\beta-\gamma)}{2}\ ,$$

Equation 3

where $$H^* = 2(c_1 c_4 + c_2 c_3)\frac{p\alpha - p\gamma+\beta+\gamma}{2}+2c_1 c_3\{p\alpha - p\beta+\beta\}+2c_2 c_4\{p\beta - p\gamma+\gamma\}$$

Previous analyses of TDT power are special cases of the current analysis

I now show that two previous analytic investigations of TDT power (Terwilliger & Ott, 1992; Risch & Merikangas, 1996) are special cases of the current analysis. These two analytic investigations and a third by Sham & Curtis (1995) as well as simulation and computer-based analyses of TDT power (Schaid & Sommer, 1994; Clerget-Darpoux *et al.* 1995; Kaplan *et al.* 1997) all assumed the disease locus to be bi-allelic. Sham & Curtis (1995) and Kaplan *et al.* (1997) considered a multi-allele marker locus, but the other analyses assumed either a bi-allelic marker or a direct test of the disease polymorphism itself, and each analysis examined TDT power for one or several specific modes of inheritance.

Terwilliger & Ott (1992) considered a recessive disease with no phenocopies in families ascertained through a single affected child. By investigating the same recessive model, Ott (1989) had previously derived an algebraic probability for transmission of each marker allele (denoted H and h) to affected offspring by heterozygous H/h parents and by both types of homozygous parent (see Ott's table II). Thus, power results for the TDT (McNemar's test in figure 3 of Terwilliger & Ott) can be derived by using Ott's table II to compute P_t^* for the recessive, zero-phenocopy model. P_t^* is derived by considering only the two probabilities in Ott's table for heterozygous H/h parents, and by dividing the transmission probability for allele H by the sum of the probabilities for allele H and allele h to yield:

$$P_t^* = \frac{(m+\delta/p)(1-m)-\theta\delta/p}{(m+\delta/p)(1-m)+m[(1-m)-\delta/p]},$$

where, substituting my notation for Ott's, m is the frequency of marker allele H (or, alternatively, my allele A), p is the frequency of disease allele D, and δ is the coefficient of disequilibrium. This expression for P_t^* is seen to be a special case of equation (3) by making the appropriate penetrance substitutions ($\alpha > 0, \beta = \gamma = 0$) into my expressions for P_t^* and H^*, and by expressing c_1, c_2, c_3 and c_4 in terms of m, p and δ according to standard expressions given above (see General algebraic model of linkage).

Risch & Merikangas (1996) compared TDT and ASP power for an intermediate mode of inheritance in which D/d penetrance (β) is a multiple (k) of d/d penetrance and $r = k^2$. For their analysis, the TDT was assumed to test the disease locus itself or a perfectly associated bi-allelic marker and, under this assumption, Risch & Merikangas found that $P_t^* = P_t = k/(1+k)$. With appropriate substitutions ($\theta = 0, \alpha = k^2\gamma, \beta = k\gamma, c_1 = p, c_4 = 1-p,\ c_2 = c_3 = 0$), equations (2) and (3) for P_t and P_t^* also simplify to $k/(1+k)$, thus agreeing that for this particular model, P_t and P_t^* are (a) identical and (b) independent of disease allele frequency. For many other genetic models, P_t and P_t^* do appear to be similar in value (though not identical). However, graphical analysis of P_t shows that the value of P_t is independent of disease allele frequency only for the particular mode of inheritance considered by Risch & Merikangas (graphs showing this are available from the author).

The expression in Risch & Merikangas (1996) for P_s (their Y) when a marker is fully informative can also be shown to be a special case of equation (1) for P_s. This is verified by substituting the appropriate mode of inheritance parameters ($\alpha = k^2\gamma, \beta = k\gamma$) into equation (1) and also by substituting parameters for a closely linked marker in equilibrium with the disease locus (i.e. $\theta = 0$ and $\delta = 0$ in $c_1 = mp+\delta,\ c_2 = m(1-p)-\delta,\ c_3 = (1-m)p-\delta$ and $c_4 = (1-m)(1-p)+\delta$). In Appendix IV and Results (see Power of χ^2_{asp} and χ^2_{tdt}), I showed that when a bi-allelic marker is in equilibrium with the disease locus, equation (1) for P_s is identical to the expression for P_s when a marker is fully informative as was assumed by Risch & Merikangas (1996).

Proportion of A/B parents in ascertained families

Since power analyses often calculate power for a specific number of ascertained families, the proportion of informative A/B parents in such families must be calculated to determine the subset of parents to which the TDT or ASP test is applied when a marker is bi-allelic. Among parents ascertained through one affected child, it can be shown that the expected proportion of A/B parents is H^*/F^* where H^* is as previously defined (see equation (3)) and $F^* = p^2\alpha+2p(1-p)\beta+(1-p)^2\gamma$. Similarly, the expected proportion of A/B parents among those ascertained through an ASP can be shown to be H/F where H is as defined in Appendix I and

$$F = p^4\alpha^2+4p^3(1-p)\ \frac{\alpha+\beta}{2}^{2} + 2p^2(1-p)^2\beta^2+4p^2(1-p)^2\ \frac{\alpha+2\beta+\gamma}{4}^{2} +4p(1-p)^3\ \frac{\beta+\gamma}{2}^{2} +(1-p)^4\gamma^2.$$

With appropriate substitutions ($\alpha = k^2\gamma, \beta = k\gamma$), the expressions H^*/F^* and H/F reduce to the corresponding expressions given by Risch & Merikangas (1996) for the proportion of heterozygous parents found in families having at least one and two affected children, respectively. Furthermore, for the recessive, zero-phenocopy disease considered by Ott (1989), the sum of the two probabilities for heterozygous parents in Ott's table II gives the proportion of heterozygous parents in families ascertained through a single affected child. When appropriate substitutions are made ($\alpha > 0, \beta = \gamma = 0$), the expression H^*/F^* also reduces to the proportion of heterozygous parents predicted by Ott's table. It is also important to note that Sham & Curtis (1995) derived a table of probabilities analogous to Ott's table II, except that their table 3 has entries for a variable number of marker alleles and their probabilities describe a general model of disease. If table 3 of Sham & Curtis is assumed to have only two marker alleles, then the two probabilities for heterozygous parents predict a probability of allele transmission identical to P_t^* (equation (3)) as well as a proportion of heterozygous parents in ascertained families which is identical to H^*/F^*.

Power of χ^2_{tdt} for a multi-allelic marker

So far the four haplotype frequencies (c_1, c_2, c_3, c_4) have represented a bi-allelic marker linked to a bi-allelic disease locus; but these frequencies could also correspond to any two marker alleles (a_i, a_j) of a multi-allelic marker linked to a bi-allelic disease locus [$c_1 = f(a_i\,\text{D}), c_2 = f(a_i\,d), c_3 =$

$f(a_j\,D), c_4 = f(a_j\,d)$]. The expression for H^*/F^* (or H/F) would then be the proportion of ascertained parents expected to be heterozygous for a_i and a_j, and the expression for P_t^* (or P_t) would be the conditional probability that a_i/a_j parents transmit allele a_i to affected offspring. Thus, in principle, the expressions for P_t^* (or P_t) and H^*/F^* (or H/F) could be used to investigate the power of any strategy for applying the TDT to a multi-allelic marker.

Here I briefly discuss a strategy recommended by Spielman & Ewens (1996) in which χ^2_{tdt} is calculated for each allele i of a multi-allelic marker ($i = 1$ to k) by evaluating parents heterozygous for allele i and the other alleles grouped together (non-i). The marker is then tested for linkage by evaluating the statistical significance of the largest of the k χ^2_{tdt}'s using significance cutpoints adjusted for multiple testing and non-independence of the chi-squares (see Ewens & Spielman [1997] for a table of these cutpoints). The power of this procedure can be estimated for any multi-allelic marker model as follows: For each i/non-i determine the haplotype frequencies c_1, c_2, c_3, c_4 and calculate the associated values of P_t^* and H^*/F^* (or P_t, H/F and P_s). Then determine the i/non-i likely to give the highest χ^2_{tdt} by calculating the expected value of each χ^2_{tdt} [$E(\chi^2_{\text{tdt}})$]. (For S parents of an ASP, it can be shown that $E(\chi^2_{\text{tdt}}) = 2(S-1)(2P_t-1)^2+2P_s$ while for singletons, $E(\chi^2_{\text{tdt}}) = (S-1)(2P_t^*-1)^2+1$.) For the i/non-i giving the highest $E(\chi^2_{\text{tdt}})$, TDT power would then be determined exactly as for a bi-allelic marker (see Power of χ^2_{tdt} and χ^2_{asp}) except that an adjusted significance cutpoint would be used as described above.

To briefly examine power for a particular multi-allelic example, consider the bi-allelic marker and disease locus in the bottom line of Table 2 ($r = 4$). The frequencies (p and m) of disease allele D and positively associated marker allele A are 0.15 and 0.25, respectively, and TDT power is 0.99 when $\delta = \delta_{\max}$ and 0.86 when $\delta = \frac{1}{2}\delta_{\max}$. Suppose p and m remain constant as does the degree of positive association between alleles A and D ($\delta = \delta_{\max}$ or $\frac{1}{2}\delta_{\max}$) but suppose the marker consists of $k-1$ additional (non-A) alleles having negative or no association with disease allele D. Then A/non-A would give the highest $E(\chi^2_{\text{tdt}})$ of any i/non-i and thus TDT power would be determined by the P_t and H/F shown in the bottom line of the table. According to Ewens & Spielman (1997), adjusted cutpoints (0.05 significance) for $k = 2$, $k = 4$ and $k = 8$ are $\chi^2_{\text{tdt}} = 3.84$, 6.10 and 7.41, respectively; thus TDT power when $k = 2$, 4 or 8 would be 0.86, 0.71 or 0.61 at $\delta = \frac{1}{2}\delta_{\max}$ and would be 0.99 for $k \leqslant 8$ when $\delta = \delta_{\max}$. This example suggests that TDT power for a multi-allelic marker remains relatively strong if (a) one marker allele is strongly associated with either allele of a bi-allelic disease locus and (b) the two associated alleles have similar population frequencies.

Concluding remarks

Strength of evidence for linkage provided by χ^2_{asp} and χ^2_{tdt} critically depends upon the magnitude of departure from the null hypothesis value of 0.5, the size of departure being quantified by $(P_s-0.5)$ and $|P_t-0.5|$ for the ASP and TDT paradigms, respectively. In this paper, I have shown that $(P_s-0.5)$ and $|P_t-0.5|$ are each a product of three corresponding factors [$(P_s-0.5) = L_s M_s R_s$, $|P_t-0.5| = L_t|M_t|R_t$]. L_s and L_t depend only on the recombination fraction (θ), R_s and R_t depend only on disease penetrance (α, β, γ) and the frequency (p) of the disease allele and, furthermore, marker allele frequency (m) and disequilibrium (δ) influence only M_s and $|M_t|$. Hence, the corresponding factors in $(P_s-0.5)$ and $|P_t-0.5|$ facilitate comparisons between the ASP and TDT paradigms, and also enable some 'partitioning' of the contribution to evidence for linkage provided by standard genetic variables such as θ, δ, m, etc.

Together with the expression for parental heterozygosity at the marker (H/F), the expressions for P_s and P_t provide a general framework for calculating and comparing the power of χ^2_{asp} and χ^2_{tdt}. This framework generalizes the ASP-TDT comparison of Risch & Merikangas (1996) by encompassing many modes of inheritance rather than just one, and also by enabling TDT power to be calculated for a marker that is distinct from the disease locus. Analysis of the equations shows that TDT power is greatly increased if disequilibrium is strong and if the disease allele and positively associated marker allele have similar population frequencies. The equations also show that the superior power of the TDT compared to the ASP test is greatest when susceptibility loci confer modest disease risk, as indicated by low values of the penetrance ratio r. When a marker is strongly associated with a disease locus that contributes modest disease risk, $|P_t - 0.5| \gg (P_s - 0.5) \approx 0$. Thus, the TDT is likely to play an important role in detecting and replicating linkages to loci responsible for complex genetic disease.

I am deeply grateful to Richard Spielman for encouragement and valuable suggestions as this work developed. I am also indebted to Warren Ewens for valuable comments and for criticism that improved the manuscript. This research was supported by NIH grants DK46618 and DK47481 and by grant 193189 from the Juvenile Diabetes Foundation.

REFERENCES

Bell, G. I., Horita, S. & Karam, J. H. (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33**, 176–183.

Blackwelder, W. C. & Elston, R. C. (1985) A comparison of sib-pair linkage tests for susceptibility loci. *Genet. Epidemiol.* **2**, 85–97.

Clerget-Darpoux, F., Babron, M. C. & Bickeböller, H. (1995) Comparing the power of linkage detection by the transmission disequilibrium test and identity-by-descent test. *Genet. Epidemiol.* **12**, 583–588.

Cox, N. J. & Spielman, R. S. (1989) The insulin gene and susceptibility to IDDM. *Genet Epidemiol* **6**, 65–69.

Ewens, W. J. & Spielman, R. S. (1997) Disease associations and the transmission/disequilibrium test (TDT). *Current Protocols in Human Genetics* 1.12.1–1.12.13.

Julier, C., Hyer, R. N., Davies, J., Merlin, F., Soularue, P., Briant, L., Cathelineau, G., *et al.* (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* **354**, 155–159.

Kaplan, N. L., Martin, E. R. & Weir, B. S. (1997) Power studies of the transmission/disequilibrium tests with multiple alleles. *Am. J. Hum. Genet.* **60**, 691–702.

McGinnis, R. E., Spielman, R. S. & Ewens, W. J. (1991) Linkage between the insulin gene (IG) region and susceptibility to insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet. Suppl.* **49**, A476.

Müller-Myshsok, B. & Abel, L. (1997) Genetic analysis of complex diseases. *Science* **275**, 1328–1329.

Ott, J. (1989) Statistical properties of the haplotype relative risk. *Genet. Epidemiol.* **6**, 127–130.

Ott, J. (ed) (1991) *Analysis of human genetic linkage.* Johns Hopkins University Press, Baltimore.

Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* **46**, 229–241.

Risch, N. & Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.

Pearson, E. S. & Hartley, H. O. (ed) (1954) *Biometrika tables for statisticians.* Vol 1. Cambridge University Press.

Schaid, D. J. & Sommer, S. S. (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am. J. Hum. Genet.* **55**, 402–409.

Sham, P. C. & Curtis, D. (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* **59**, 323–336.

Spielman, R. S., Baur, M. P. & Clerget-Darpoux, F. (1989). Genetic analysis of IDDM: summary of GAW5 IDDM results. *Genet. Epidemiol.* **6**, 43–58.

Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516.

Spielman, R. S. & Ewens, W. J. (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**, 983–989.

Terwilliger, J. D. & Ott, J. (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346.

Thomson, G., Robinson, W. P., Kuhner, M. K. & Joe, S. (1989) HLA, insulin gene, and Gm associations with IDDM. *Genet. Epidemiol.* **6**, 155–160.

Weir, B. S. (ed) (1996) *Genetic data analysis II*, 2nd ed., Sinauer Associates Inc., Sunderland, MA.

APPENDIX I

Derivation of expressions for P_s, P_t, H

The derivations assume the general model of a bi-allelic marker and linked bi-allelic disease locus that is the only locus that underlies disease susceptibility (see General algebraic model of linkage in the main text). I begin the derivation of P_s and P_t (equations (1) and (2)) by first deriving

Table A1. *Conditional on mating type, the probability that allele D or d is transmitted from a specific D/d parent to any specific affected child*

	Probability of transmission to affected	
Mating type	Allele D	Allele d
$D/d \times D/D$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\beta}{\alpha+\beta}$
FCI D/d × D/d	$\frac{\alpha+\beta}{\alpha+2\beta+\gamma}$	$\frac{\beta+\gamma}{\alpha+2\beta+\gamma}$
$D/d \times d/d$	$\frac{\beta}{\beta+\gamma}$	$\frac{\gamma}{\beta+\gamma}$

Table A2. *Families with a heterozygous A/B father and N children, at least k of whom are affected*

Mating type		Probability of transmission to affected	
Fa Mo	'Weighted' frequency	Allele A	Allele B
$AD/Bd \times D/D$	$2(c_1 c_4)p^2N \frac{\alpha+\beta}{2}^{k}$	$\frac{\alpha-\theta(\alpha-\beta)}{\alpha+\beta}$	$\frac{\beta-\theta(\beta-\alpha)}{\alpha+\beta}$
$AD/Bd \times D/d$	$4(c_1 c_4)p(1-p)N \frac{\alpha+2\beta+\gamma}{4}^{k}$	$\frac{(\alpha+\beta)-\theta(\alpha-\gamma)}{\alpha+2\beta+\gamma}$	$\frac{(\beta+\gamma)-\theta(\gamma-\alpha)}{\alpha+2\beta+\gamma}$
$AD/Bd \times d/d$	$2(c_1 c_4)(1-p)^2N \frac{\beta+\gamma}{2}^{k}$	$\frac{\beta-\theta(\beta-\gamma)}{\beta+\gamma}$	$\frac{\gamma-\theta(\gamma-\beta)}{\beta+\gamma}$
$Ad/BD \times D/D$	$2(c_2 c_3)p^2N \frac{\alpha+\beta}{2}^{k}$	$\frac{\beta-\theta(\beta-\alpha)}{\alpha+\beta}$	$\frac{\alpha-\theta(\alpha-\beta)}{\alpha+\beta}$
$Ad/BD \times D/d$	$4(c_2 c_3)p(1-p)N \frac{\alpha+2\beta+\gamma}{4}^{k}$	$\frac{(\beta+\gamma)-\theta(\gamma-\alpha)}{\alpha+2\beta+\gamma}$	$\frac{(\alpha+\beta)-\theta(\alpha-\gamma)}{\alpha+2\beta+\gamma}$
$Ad/BD \times d/d$	$2(c_2 c_3)(1-p)^2N \frac{\beta+\gamma}{2}^{k}$	$\frac{\gamma-\theta(\gamma-\beta)}{\beta+\gamma}$	$\frac{\beta-\theta(\beta-\gamma)}{\beta+\gamma}$
$AD/BD \times D/D$	$2(c_1 c_3)p^2N\alpha^k$	0.5	0.5
$AD/BD \times D/d$	$4(c_1 c_3)p(1-p)N \frac{\alpha+\beta}{2}^{k}$	0.5	0.5
$AD/BD \times d/d$	$2(c_1 c_3)(1-p)^2N\beta^k$	0.5	0.5
$Ad/Bd \times D/D$	$2(c_2 c_4)p^2N\beta^k$	0.5	0.5
$Ad/Bd \times D/d$	$4(c_2 c_4)p(1-p)N \frac{\beta+\gamma}{2}^{k}$	0.5	0.5
$Ad/Bd \times d/d$	$2(c_2 c_4)(1-p)^2N\gamma^k$	0.5	0.5

expressions for two related probabilities denoted P_{2A} and P_{2B}. In the main text, I assume ascertainment of parents through a randomly selected ASP and note that equations (1) and (2) apply to an ascertained parent who is also informative A/B at a bi-allelic marker. P_{2A} is the probability that the parent transmitted allele A to both offspring in the ASP, and P_{2B} is the probability that B was transmitted to both. Since P_s is the probability that an ascertained A/B parent transmitted the same marker allele to an ASP, it follows that $P_s = P_{2A} + P_{2B}$. Similarly, P_t is the probability that the A/B parent transmitted allele A to a specific affected child, and hence

$$P_t = P_{2A} + \tfrac{1}{2}(1 - P_{2A} - P_{2B}) = \tfrac{1}{2} + \tfrac{1}{2}(P_{2A} - P_{2B}).$$

Therefore, the expressions for P_s and P_t can be found by deriving expressions for $(P_{2A} + P_{2B})$ and for $(P_{2A} - P_{2B})$. As a preliminary step in finding these expressions, I present Table A1 which shows the three mating types that contain at least one parent who is heterozygous (D/d) at the disease locus. Conditional upon mating type, the table gives the probability that a particular D/d parent in the family transmitted allele D or allele d to any specific affected child. Assuming that the bi-allelic disease locus is the only locus responsible for disease susceptibility, it can be shown that the conditional probabilities in Table A1 are valid, regardless of the number of children in the family or the number who are affected.

Using the probabilities in Table A1, I now derive expressions for P_{2A} and P_{2B}. To simplify the derivation, I calculate P_{2A} (P_{2B}) as the conditional probability that allele A (allele B) was transmitted to two affected sibs by an *A/B father* randomly selected from a subpopulation of families having exactly N children, two or more of whom are affected. I then show that the results are identical to those for *A/B parents* randomly selected from families of *any size* that have two or more affected offspring.

Based in part on the conditional probabilities in Table A1, the entries in Table A2 describe families with an A/B father and N children, at least k of whom are affected. For a subpopulation of families having exactly N children, the subpopulation frequency for each mating type in column 1 of Table A2 can be subdivided into a summation of component frequencies with each component corresponding to the exact number (r) of affected children in the family. Specifically, the component frequency for each mating type with exactly r affected children ($k \leqslant r \leqslant N$) is found by multiplying three factors: (a) the general population frequency of the father's genotype as defined by the two marker-disease locus haplotypes in the A/B father times (b) the general population frequency of the disease locus genotype in the mother times (c) the conditional probability that the mating type produces exactly r affected offspring among the N children. For the mating type on the first line of Table A2, the subpopulation or component frequency for families with r affected offspring would be:

$$2(c_1 c_4)p^2 \; {N \atop r} \; \frac{\alpha+\beta}{2}^{\,r} \; 1-\frac{\alpha+\beta}{2}^{\,N-r}.$$

The probability that a particular mating type with r affected children will be randomly selected (ascertained) from the subpopulation is directly proportional to the component frequency 'weighted' by (multiplied by) a factor (f_{w}) that depends on the random ascertainment scheme. When only one affected child is required for family ascertainment ($k = 1$), the probability of ascertainment is directly proportional to the number (r) of affected children in the family since any of these is a potential proband. Thus $f_{\mathrm{w}} = r$ when $k = 1$. When an affected sib pair ($k = 2$) is required for ascertainment, I adopt the ascertainment scheme for ASPs that is implicit in the calculation of λ_{s} (Risch, 1990). In this scheme, a proband (affected individual) is randomly selected from the population and then one of the proband's sibs (affected or unaffected) is randomly selected. Only pairs in which the second sib is affected contribute to the magnitude of the recurrence risk (K_{s}) and to the magnitude of $\lambda_{\mathrm{s}} = K_{\mathrm{s}}/K$ (see Risch, 1990) and thus only these pairs (and the families that produced them) are ascertained. By contrast, pairs in which the second sib is unaffected do not contribute to the magnitude of K_{s} or λ_{s} and are, in effect, discarded (family not ascertained). Based on this ascertainment scheme for ASPs, the component frequency for a mating type with r affected offspring would be weighted by $f_{\mathrm{w}} = r(r-1)/(N-1)$ since $(r-1)/(N-1)$ is the probability of randomly selecting a second affected sib from the $N-1$ children who remain after proband selection.

For each mating type in column 1 of Table A2, the probability of randomly selecting or ascertaining a family with at least k affected offspring is directly proportional to the sum of the weighted component frequencies for which r is greater than or equal to k (i.e. $k \leqslant r \leqslant N$). For each mating type in Table A2, this sum simplifies to the overall 'weighted' subpopulation frequency shown in column 2. For example, when $k = 2$ (and hence $f_w = r(r-1)/(N-1)$), the sum for the mating type in line 1 simplifies as follows:

$$\sum_{r=2}^{N} r\frac{r-1}{N-1}2(c_1 c_4)p^2 \genfrac{}{}{0pt}{}{N}{r} \;{\frac{\alpha+\beta}{2}}^{r} \; 1-{\frac{\alpha+\beta}{2}}^{N-r}$$

$$= 2(c_1 c_4)p^2N \;{\frac{\alpha+\beta}{2}}^{2} \sum_{r=2}^{N} \genfrac{}{}{0pt}{}{N-2}{r-2} \;{\frac{\alpha+\beta}{2}}^{r-2} \; 1-{\frac{\alpha+\beta}{2}}^{N-r}$$

$$= 2(c_1 c_4)p^2N \;{\frac{\alpha+\beta}{2}}^{2} \sum_{r=0}^{N-2} \genfrac{}{}{0pt}{}{N-2}{r} \;{\frac{\alpha+\beta}{2}}^{r} \; 1-{\frac{\alpha+\beta}{2}}^{N-2-r}$$

$$= 2(c_1 c_4)p^2N \;{\frac{\alpha+\beta}{2}}^{2} \;\frac{\alpha+\beta}{2}+1-{\frac{\alpha+\beta}{2}}^{N-2}$$

$$= 2(c_1 c_4)p^2N \;{\frac{\alpha+\beta}{2}}^{2}.$$

Based on the 'weighted' subpopulation frequencies in column 2, what then is the probability of randomly selecting a particular 'A/B father-mating type' from among families with an A/B father and N children, at least two of whom are affected? Setting $k = 2$, the probability (P_{rs}) of randomly selecting a particular mating type would be the weighted frequency of the mating type divided by the sum of all the weighted frequencies in column 2. Thus, the probability of randomly selecting the mating type on line 1 ($AD/Bd \times D/D$) would be:

$$P_{\text{rs}} = \frac{2(c_1 c_4)}{H}p^2 \;{\frac{\alpha+\beta}{2}}^{2},$$

where

$$H = 2(c_1 c_4 + c_2 c_3)\;\; p^2 \;{\frac{\alpha+\beta}{2}}^{2} + \tfrac{1}{2}p(1-p)\;{\frac{\alpha+2\beta+\gamma}{2}}^{2} + (1-p)^2\;{\frac{\beta+\gamma}{2}}^{2}$$
$$+2c_1 c_3\{p^2\alpha^2 + \tfrac{1}{2}p(1-p)(\alpha+\beta)^2 + (1-p)^2\beta^2\}$$
$$+2c_2 c_4\{p^2\beta^2 + \tfrac{1}{2}p(1-p)(\beta+\gamma)^2 + (1-p)^2\gamma^2\}.$$

Note that the quantity N cancels in the numerator and denominator of P_{rs}, thus demonstrating that the probability is independent of family size (N) and hence applies to mixed populations of families of *any* size that have 2 or more affected offspring. Furthermore, if mating types with A/B mothers were included in Table A2, the only effect would be to double each frequency in column 2; but P_{rs} would be unchanged since the '2's' in the doubled numerator and denominator of P_{rs} would cancel. Thus, by setting $k = 2$, the P_{rs} calculated for each mating type in Table A2 applies to random selection of A/B *parents* of two or more affected offspring from families of *any* size.

What, then, is the probability that a randomly selected A/B parent of a particular mating type transmitted allele A (allele B) to an affected child? For each mating type in Table A2, the two rightmost columns show the conditional probability that the A/B parent transmitted allele A or B to an individual affected offspring. These probabilities follow directly from the conditional probabilities in Table A1. For example, in the $AD/Bd \times D/D$ mating in line 1 of Table A2, the A/B parent has allele A in coupling with allele D, and B in coupling with d, and thus Table A1 implies that A is transmitted to affected offspring with a probability of

$$(1-\theta)\frac{\alpha}{\alpha+\beta}+\theta\frac{\beta}{\alpha+\beta} = \frac{\alpha-\theta(\alpha-\beta)}{\alpha+\beta}.$$

Therefore, from the results of the previous paragraph, the joint probability that (a) the $AD/Bd \times D/D$ mating type (line 1) is randomly selected and (b) the A/B parent transmitted allele A to both affected sibs of an ASP would be:

$$\frac{2(c_1c_4)p^2}{H}\,\frac{\alpha+\beta}{2}^{2}\cdot\,\frac{\alpha-\theta(\alpha-\beta)}{\alpha+\beta}^{2}=\frac{2(c_1c_4)p^2}{H}\,\frac{\alpha-\theta(\alpha-\beta)}{2}^{2}.$$

P_{2A} would then be the sum of this joint probability and the corresponding joint probabilities for the other 11 mating types in Table A2:

$$\begin{aligned}P_{2A}=&\frac{2(c_1c_4)}{H}\;p^2\,\frac{\alpha-\theta(\alpha-\beta)}{2}^{2}+2p(1-p)\,\frac{(\alpha+\beta)-\theta(\alpha-\gamma)}{4}^{2}+(1-p)^2\,\frac{\beta-\theta(\beta-\gamma)}{2}^{2}\\&+\frac{2(c_2c_3)}{H}\;p^2\,\frac{\beta-\theta(\beta-\alpha)}{2}^{2}+2p(1-p)\,\frac{(\beta+\gamma)-\theta(\gamma-\alpha)}{4}^{2}+(1-p)^2\,\frac{\gamma-\theta(\gamma-\beta)}{2}^{2}\\&+\frac{2(c_1c_3)}{H}\;p^2\,\frac{\alpha}{2}^{2}+2p(1-p)\,\frac{\alpha+\beta}{4}^{2}+(1-p)^2\,\frac{\beta}{2}^{2}\\&+\frac{2(c_2c_4)}{H}\;p^2\,\frac{\beta}{2}^{2}+2p(1-p)\,\frac{\beta+\gamma}{4}^{2}+(1-p)^2\,\frac{\gamma}{2}^{2}\end{aligned}$$

Similarly, P_{2B} equals:

$$\begin{aligned}P_{2B}=&\frac{2(c_1c_4)}{H}\;p^2\,\frac{\beta-\theta(\beta-\alpha)}{2}^{2}+2p(1-p)\,\frac{(\beta+\gamma)-\theta(\gamma-\alpha)}{4}^{2}+(1-p)^2\,\frac{\gamma-\theta(\gamma-\beta)}{2}^{2}\\&+\frac{2(c_2c_3)}{H}\;p^2\,\frac{\alpha-\theta(\alpha-\beta)}{2}^{2}+2p(1-p)\,\frac{(\alpha+\beta)-\theta(\alpha-\gamma)}{4}^{2}+(1-p)^2\,\frac{\beta-\theta(\beta-\gamma)}{2}^{2}\\&+\frac{2(c_1c_3)}{H}\;p^2\,\frac{\alpha}{2}^{2}+2p(1-p)\,\frac{\alpha+\beta}{4}^{2}+(1-p)^2\,\frac{\beta}{2}^{2}\\&+\frac{2(c_2c_4)}{H}\;p^2\,\frac{\beta}{2}^{2}+2p(1-p)\,\frac{\beta+\gamma}{4}^{2}+(1-p)^2\,\frac{\gamma}{2}^{2}\end{aligned}$$

Therefore,

$$P_{2A}-P_{2B}=(1-2\theta)\;\frac{2c_1c_4-2c_2c_3}{H}\quad p^2\,\frac{\alpha^2-\beta^2}{4}+2p(1-p)\,\frac{(\alpha+\beta)^2-(\beta+\gamma)^2}{16}+(1-p)^2\,\frac{\beta^2-\gamma^2}{4}$$

Equation 4

Hence, equation (2) in the main text follows from the equation $P_t=\frac{1}{2}+\frac{1}{2}(P_{2A}-P_{2B})$ given at the beginning of Appendix I. Similarly, by adding P_{2A} and $P_{2B}, P_s=(P_{2A}+P_{2B})$ simplifies, after some algebra, to equation (1).

Derivation of expressions for P_t^* *and* H^*

To simplify this derivation, I use 'P_A' to denote what appears in the main text as the probability 'P_t^*'. P_A (P_B) is the probability that allele A (allele B) was transmitted to an individual affected child by a randomly ascertained A/B parent of one or more affected offspring. So by setting $k=1$ in Table A2, the derivations of P_A, P_B and H^* are analogous to the derivations of P_{2A}, P_{2B} and H described above. Thus, summing the frequencies of the mating types in Table A2 for $k=1$ and factoring out N, I obtain:

$$H^*=2(c_1c_4+c_2c_3)\frac{p\alpha-p\gamma+\beta+\gamma}{2}+2c_1c_3\{p\alpha-p\beta+\beta\}+2c_2c_4\{p\beta-p\gamma+\gamma\}$$

Similarly, P_A (P_B) equals an expression identical to the expression for P_{2A} (P_{2B}) shown above except that H^* replaces H and each quantity inside square brackets is *not* squared in the corresponding expression for P_A (P_B). Based on these expressions for P_A and P_B, the expression for $(P_A - P_B)$ simplifies in a manner analogous to $(P_{2A} - P_{2B})$ [see equation (4)]. Therefore, the expression for P_A can also be simplified by using the relation $(P_A + P_B) = 1$ which implies that $P_A = \frac{1}{2} + \frac{1}{2}(P_A - P_B)$. This simplified expression for P_A is shown in the main text (equation (3) for P_t^*).

APPENDIX II

Power of χ^2_{asp} and χ^2_{tdt}

Assume that each of S parents is independently and randomly ascertained through an ASP. Suppose we wish to determine the power of χ^2_{asp} and χ^2_{tdt} when applied to those parents who are informative (A/B) at a bi-allelic marker. If there are h heterozygous A/B parents, and if χ^2_{asp} and χ^2_{tdt} consider one ASP per parent, then the random variables i and j denote the following three subdivisions of the h parents:

i = number of A/B parents who transmit allele A to both affected offspring, or the number of ASPs that share allele A

j = number of A/B parents who transmit allele B to both affected offspring, or the number of ASPs that share allele B

$(h-i-j)$ = number of A/B parents who transmit allele A to one affected offspring and allele B to the other, or the number of ASPs that share neither allele A nor B

As shown in Spielman *et al.* (1993), χ^2_{tdt} and χ^2_{asp} can be expressed in terms of h, i and j as follows: $\chi^2_{\text{tdt}} = [2(i-j)^2]/h$ and $\chi^2_{\text{asp}} = [(2(i+j)-h)^2]/h$. Hence, the probability of each possible (i,j) combination gives the probability of each possible value of χ^2_{tdt} and χ^2_{asp}.

What, then, is the probability of each (i,j) combination, when a total of h A/B parents are randomly selected (ascertained) from A/B parents of one or more ASPs? The probability that the first randomly selected parent transmitted A or B to both affected offspring is given by P_{2A} and P_{2B}, respectively (see Appendix I). Furthermore, sampling with replacement would apply, if h is small relative to the population being sampled. Therefore, P_{2A}, P_{2B} and $(1-P_{2A}-P_{2B})$ would be the probabilities of the three possible outcomes generated by each randomly selected parent, and since the random selections are independent, the joint probability distribution for (i,j) is:

$$P(i,j) = \frac{h!}{i!j!(h-i-j)!}(P_{2A})^i (P_{2B})^j (1-P_{2A}-P_{2B})^{h-i-j}. \qquad \text{Equation 5}$$

Consequently, χ^2_{asp} power or the probability that $\chi^2_{\text{asp}} > L$ (a significance cutpoint) is given by the sum of the $P(i,j)$ terms for which $(i+j) > h/2 + \sqrt{(hL)}/2$. Similarly, the power of χ^2_{tdt} is the sum of $P(i,j)$ terms for which $|i-j| > \sqrt{(hL/2)}$. However, these power determinations are simplified numerically and conceptually by reducing the trinomial distribution to one of two binomial distributions. With respect to χ^2_{asp}, h randomly selected A/B parents can be regarded as a series of h binomial trials, each having $(P_{2A}+P_{2B})$ probability of generating an ASP that shares parental allele A or B. Consequently, χ^2_{asp} power is determined by the binomial distribution

$$\frac{h!}{(i+j)!\,(h-i-j)!}(P_{2A}+P_{2B})^{i+j}(1-P_{2A}-P_{2B})^{h-i-j}$$

or, substituting notation from the main text,

$$\frac{n_{\text{asp}}!}{n_{\text{s}}!n_{\text{u}}!}P_{\text{s}}^{n_{\text{s}}}(1-P_{\text{s}})^{n_{\text{u}}}$$

with power specifically equal to the portion of the distribution for which

$$n_{\text{s}} > n_{\text{asp}}/2+\surd(n_{\text{asp}}L)/2.$$

Although the binomial distribution based on P_{s} and n_{asp} trials enables exact calculation of χ^2_{asp} power, the power of χ^2_{tdt} can be precisely estimated, but not calculated exactly, from a second binomial distribution generated by P_{t}. Among h randomly selected A/B parents of an ASP, each parent would transmit allele A to an individual affected child with a probability of $P_{2A}+\frac{1}{2}(1-P_{2A}-P_{2B}) = \frac{1}{2}+\frac{1}{2}(P_{2A}-P_{2B}) = P_{\text{t}}$. This suggests that χ^2_{tdt} power can be determined by the binomial distribution

$$\frac{n_{\text{tdt}}!}{n_{\text{a}}!n_{\text{b}}!}P_{\text{t}}^{n_{\text{a}}}(1-P_{\text{t}})^{n_{\text{b}}} \quad \text{where} \quad n_{\text{tdt}} = 2h, \quad n_{\text{a}} = h+(i-j), \quad \text{and} \quad n_{\text{b}} = h-(i-j).$$

The estimated power of χ^2_{tdt} would then equal the portion of the distribution for which $|n_{\text{a}}-n_{\text{b}}| > \surd(n_{\text{tdt}}L)$. This is equivalent to summing the two 'tails' of the distribution, i.e. the portion of the distribution for which $n_{\text{a}} > n_{\text{tdt}}/2+\surd(n_{\text{tdt}}L)/2$ plus the portion for which $n_{\text{b}} > n_{\text{tdt}}/2+\surd(n_{\text{tdt}}L)/2$.

Thus, computer summation of the binomial terms in the two tails gives an estimate of χ^2_{tdt} power that can be compared with exact power obtained by summing the terms of the $P(i,j)$ trinomial distribution (equation (5)) for which $|i-j| > \surd(hL/2)$. I performed this comparison by calculating P_{t}, P_{2A} and P_{2B} for 1040 genetic parameter combinations that derived from five different groups (209 combinations per group). For three groups, the marker and disease locus were assumed to be identical, but the groups differed by assuming an $\alpha{:}\gamma$ penetrance ratio (r) of 2, 4 or ∞. Mode of inheritance (x) for these groups was defined by the numerical 'distance' of D/d penetrance (β) between D/D penetrance ($\alpha = r\gamma$) and d/d penetrance (γ), i.e. $\beta = x(r-1)\gamma+\gamma$. By allowing x to vary between $x = 0$ and $x = 1$ in increments of 0.1 and allowing disease allele frequency (p) to vary between $p = 0.05$ and $p = 0.95$ in increments of 0.05, 209 ordered (x,p) pairs or parameter combinations were formed for each of the three groups ($11\times 19 = 209$). The two remaining groups were for $r = 4$ or $r = \infty$, and both groups assumed a bi-allelic marker with equally frequent alleles ($m = 0.5$) to be tightly linked ($\theta = 0$) to a bi-allelic disease locus with additive mode of inheritance ($\beta = (\alpha+\gamma)/2$). For these two groups, the variable x represented degree of disequilibrium between marker and disease locus ($x = 0$ was equilibrium, $x = 1$ maximum disequilibrium). 209 parameter pairs were formed for different combinations of disequilibrium (x) and disease allele frequency (p).

By assuming a data set of 100 A/B parents of an ASP, I calculated χ^2_{tdt} power at each parameter combination for significance levels of 0.05, 0.01 and 0.001 by obtaining (a) the binomial power estimate based on P_{t} and (b) the exact trinomial power value based on P_{2A} and P_{2B}. The exact trinomial value and binomial estimate differed by less than 0.01 for 95% of the 3135 possibilities tested and by less than 0.02 for 98%. The largest overestimate of TDT power differed from the exact value by 0.037 and the largest underestimate differed by 0.052. These results show that the binomial estimate based on P_{t} is very precise.

APPENDIX III

Influence of disequilibrium on P_t and P_s

To demonstrate how disequilibrium affects P_t and P_s, I assume all parameters describing a bi-allelic marker and linked, bi-allelic disease locus are fixed, except for the degree of disequilibrium (δ) between the two loci. Hence, all parameters are constant in P_s (equation (1)) and P_t (equation (2)) except for the haplotype frequencies (c_1, c_2, c_3, c_4) in the middle factor (M_s or M_t) of each equation $[c_1 = mp+\delta, c_2 = m(1-p)-\delta, c_3 = (1-m)p-\delta, c_4 = (1-m)(1-p)+\delta]$. Therefore, by rewriting M_s and M_t in terms of δ, m and p, and then determining the δ value that maximizes M_s (or $|M_t|$) we also determine the δ that maximizes ($P_s-0.5$) (or $|P_t-0.5|$).

I focus first on M_s, the middle factor in P_s. By definition, $M_s = (c_1 c_4 + c_2 c_3)/H$, where $H = 2(c_1 c_4 + c_2 c_3) W_{Dd} + 2c_1 c_3 W_{DD} + 2c_2 c_4 W_{dd}$, and W_{Dd}, W_{DD} and W_{dd} are functions of p, α, β and γ, and hence are constant (see Appendix I for full expression for H). The δ values that minimize and maximize M_s are found by solving the equation $\partial M_s/\partial\delta = 0$ for δ. There are two solutions (roots) but only one root (δ_0) falls with the interval of genetically possible values of δ bounded by $\delta_{\min}$ and $\delta_{\max}$:

$$\delta_0 = \frac{2m(1-m)Y - \surd[m(1-m)]\cdot\surd[(W_{DD})(W_{dd})(4m(1-m)-1)+Y^2]}{(2m-1)(W_{DD}-W_{dd})}$$

where $Y = [p(W_{DD}-W_{dd})+W_{dd}]$.

Since it can be shown that ($\partial^2 M_s/\partial\delta^2 > 0$ when $\delta = \delta_0$, it follows that $\delta = \delta_0$ minimizes M_s, and P_s. Because there are no other maxima or minima for M_s between $\delta_{\min}$ and $\delta_{\max}$, the δ value that maximizes M_s, and P_s must be one of the two endpoints of the interval ($\delta_{\min}$ or $\delta_{\max}$). This is true even when $\delta = \delta_0$ happens not to fall within the interval since $\partial M_s/\partial\delta$ would then be positive or negative throughout the interval, implying that M_s is maximized at one endpoint and minimized at the other.

I now turn to M_t, and show that $|P_t-0.5|$ is always maximized at the same value ($\delta_{\min}$ or $\delta_{\max}$) that maximizes ($P_s-0.5$). By definition, $M_t = (c_1 c_4 - c_2 c_3)/H = \delta/H$ where H is as defined above. Since $H > 0$, $M_t = \delta/H$ may be positive or negative. Therefore, when all parameters in P_t are fixed except δ, the maximum value of the function $|M_t(\delta)|$ also maximizes $|P_t-0.5|$. In this connection, note that $M_t = (c_1 c_4 - c_2 c_3)/H$ is identical to $M_s = (c_1 c_4 + c_2 c_3)/H$ except that in the numerator, $c_1 c_4$ and $c_2 c_3$ are added rather than subtracted. Since H and the haplotype frequencies c_1, c_2, c_3 and c_4 are always positive, $M_s \geqslant |M_t|$ at any value of δ. However, it is well known that $\delta_{\max} = m(1-p)$ if $p \leqslant m$ or $\delta_{\max} = p(1-m)$ if $p > m$, and that $\delta_{\min} = -mp$ if $p \leqslant (1-m)$ or $\delta_{\min} = -(1-m)(1-p)$ if $p > (1-m)$ [Ott, 1991]. Since $c_1 = mp+\delta$, $c_2 = m(1-p)-\delta$, $c_3 = (1-m)p-\delta$ and $c_4 = (1-m)(1-p)+\delta$, the expressions for $\delta_{\max}$ and $\delta_{\min}$ imply that $c_1 c_4 = 0$ or $c_2 c_3 = 0$ at both $\delta = \delta_{\max}$ and $\delta = \delta_{\min}$. Hence, $M_s = |M_t|$ at $\delta = \delta_{\max}$ and at $\delta = \delta_{\min}$. Therefore, the same value ($\delta_{\max}$ or $\delta_{\min}$) that maximizes M_s must also maximize $|M_t|$ and $|P_t-0.5|$.

APPENDIX IV

Derivation of expression for P_s for a fully informative marker

To derive P_s for a fully informative or completely polymorphic marker, suppose that a bi-allelic marker is perfectly associated with a linked ($\theta < 0.5$), bi-allelic disease locus which implies that $c_2 = c_3 = 0, c_1 = p, c_4 = 1-p$ and hence, from equation (1), $P_s = 0.5+(1-2\theta)^2[p(1-p)/H]R_s$. As explained in the Discussion ('Proportion of A/B parents in ascertained families'), this value of P_s applies only to the proportion (H/F) of ascertained parents who are informative (A/B) at the bi-allelic marker. But, in this case, the marker and disease locus are perfectly associated, and hence these ascertained A/B parents must also be informative (D/d) at the disease locus. By contrast, because the marker and disease locus are perfectly associated, the remaining proportion $(1-H/F)$ of ascertained parents must be homozygous (D/D or d/d) at the disease locus as well as homozygous at the bi-allelic marker; hence these doubly homozygous parents would transmit each marker allele (ibd) to affected offspring with equal probability. Therefore, P_s would equal 0.5 for these doubly homozygous parents if they were made informative by testing at a linked marker that is completely polymorphic. Hence, if all parents were tested at the completely polymorphic marker, the value of P_s for the completely polymorphic marker would equal:

$$1-\frac{H}{F}\ (0.5)+\ \frac{H}{F}\quad 0.5+(1-2\theta)^2\ \frac{p(1-p)}{H}\ R_s\ = 0.5+(1-2\theta)^2\ \frac{p(1-p)}{F}\ R_s.$$

As noted in Results (see Power of χ^2_{asp} and χ^2_{tdt}), this expression for P_s is identical to that obtained for a bi-allelic marker in equilibrium with a bi-allelic disease locus.